



## BIROn - Birkbeck Institutional Research Online

Gazzola, M. and Templin, T. and Mcentee-Atalianis, Lisa (2019) Measuring diversity in multilingual communication. *Social Indicators Research* , ISSN 0303-8300. (In Press)

Downloaded from: <http://eprints.bbk.ac.uk/28182/>

*Usage Guidelines:*

Please refer to usage guidelines at <http://eprints.bbk.ac.uk/policies.html> or alternatively contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).



# Measuring Diversity in Multilingual Communication

Michele Gazzola<sup>1,2</sup> · Torsten Templin<sup>2</sup> · Lisa J. McEntee-Atalianis<sup>3</sup>

Accepted: 13 July 2019  
© The Author(s) 2019

## Abstract

This article develops new indices to measure linguistic diversity. It is new in two respects: firstly, existing indices to measure the probability that in a given multilingual context communication among people speaking different languages can successfully occur are based on the assumption that communication is possible only if at least one single language is shared. This study develops new indices that describe the probability that people with different linguistic repertoires can effectively communicate not only through one common language, but also by relying on their receptive competence in multiple languages, or a mix between the two communication strategies. Secondly, it develops indices to measure the degree of diversity of language policies aimed at providing multilingual communication (through translation and interpretation). The focus, therefore, is on the organisation as collective actors rather than individuals. The indices may be relevant to the study of the political and economic implications of linguistic diversity in multilingual countries, and in the management of diversity in multilingual organisations.

**Keywords** Linguistic diversity · Indicators · Language policy · Multilingual communication · Probability theory · Public administration

---

✉ Michele Gazzola  
m.gazzola@ulster.ac.uk

Torsten Templin  
templin@math.hu-berlin.de

Lisa J. McEntee-Atalianis  
l.mcentee-atalianis@bbk.ac.uk

<sup>1</sup> Faculty of Arts, Humanities and Social Sciences, School of Applied Social and Policy Sciences, Ulster University, Jordanstown, UK

<sup>2</sup> Research Group “Economics and Language” (REAL), Faculty of Humanities and Social Sciences, Humboldt-Universität zu Berlin, Berlin, Germany

<sup>3</sup> Department of Applied Linguistics and Communication, Birkbeck, University of London, London, UK

## 1 Introduction

This article develops new indices to measure linguistic diversity. A first set of indices describe the probability that people with different linguistic repertoires can effectively communicate not only through one common language, as is often assumed in the literature, but also by relying on their receptive competence in multiple languages, or a mix between the two communication models. In addition, it develops new indices to measure the degree of diversity of language policies aimed at providing multilingual communication through translation and interpretation in linguistically diverse organisations. This article, therefore, responds to recent calls (see Sect. 4 below) by multilingual countries and organisations to better describe multilingual contexts and to improve the set of indicators available in language policy design and evaluation.

The measurement of diversity is a branch of probability theory that has been applied to many fields, including *inter alia* ecology, linguistics, physics, economics, technology and the political sciences. Diversity is defined according to three basic properties (see Stirling 2007): variety, balance, and disparity. *Variety* (or richness) is the number of categories into which system elements are apportioned, for example, the number of species in an ecological niche, or the number of official languages in a country. *Balance* (or evenness) is a function of the pattern of distribution of elements across categories, that is, it is a measurement of proportions of different types with respect to the total, for example the percentage of the population speaking each official language of a country. Finally, *disparity* (or distance) refers to the degree to which the elements of a system may be distinguished. In biology, this is interpreted as the genomic distance between species or the number of nodes separating species on a genealogical tree. In linguistics, the disparity of languages is interpreted in terms of distance between languages  $M$  and  $N$ , measured through various methods such as lexicostatistical distance or distances based on linguistic trees.<sup>1</sup>

Statistical measurements of diversity were applied to languages in a seminal paper published by Greenberg (1956), later expanded by Lieberman (1964). Greenberg presents different quantifiable indicators (*/indices*) to measure linguistic diversity. The most referenced is Greenberg's "A" index, which he calls the "monolingual non-weighted method" (also referred to as the fractionalisation index). This indicator, discussed in more depth below, is defined as the probability that an individual randomly selected in a given population does not share the same language with another randomly selected member of the same population, assuming that all individuals are monolingual. If everyone speaks the same language in the population, the value of the index is 0, if everyone speaks a different language the value is 1: the higher the index, the higher the degree of linguistic diversity in the population. The Greenberg "B" index ("monolingual weighted method") is a more general case in which the A index is weighted for each pair of languages by a factor that reflects linguistic resemblance among such languages. The B index, therefore, combines disparity with balance.

Greenberg's A and B indices are useful to describe diversity in a given linguistic environment such as a country or a region, and to explore the correlations between linguistic diversity and different socio-economic variables. The A index (and to a lesser extent the B index) have been used by economists and political scientists to explore whether linguistic diversity influences various political and socio-economic outcomes such as democratic

<sup>1</sup> For a survey see Ginsburgh and Weber (2016a: 141–154, 2016b: 109–113).

participation, growth, social cohesion, economic development, inequality, health and the provision of collective goods in a country (see Alesina et al. 1999; Laitin 2000; Fearon 2003; Desmet et al. 2009; Bossert et al. 2011). Ethno-linguistic diversity is used in cross-country analyses as an explanatory variable of development, intra-community solidarity, conflict, income distribution and health (Baldwin and Huber 2010; Esteban et al. 2012; Sturm and De Haan 2015; Laitin and Ramachandran 2016; Churchill et al. 2017). Other authors use it in a single country or region to study the relationship between ethno-linguistic diversity on the one hand, and growth and social cohesion on the other hand (van Staveren and Pervaiz 2017; Ishizawa and Stevens 2007; Fedderke et al. 2008; Schaeffer 2013). Fractionalisation indices have also been used to assess the success of UN peace-keeping missions supported by soldiers from a variety of donor countries (Bove and Ruggeri 2016).

The first shortcoming of Greenberg's indices and its derivatives, however, is the assumption that speakers are monolingual or that communication is only possible if a common language is shared. This assumption is too restrictive. Various non-mutually exclusive approaches to the communication challenge in linguistically diverse societies are available: communication in a single language (*lingua franca*), of course, but also communication based on speakers' receptive skills, translation of documents into one or several other languages, and interpretation.

The second (related) shortcoming of Greenberg's indices is that they are not applicable to an examination of linguistic diversity in multilingual organisations such as the parliament of a multilingual country or the general assembly of an international organisation. To guarantee the functioning of the public administration in officially multilingual countries (e.g. Switzerland, Canada, and South Africa), and in international organisations such as the European Union or the United Nations, individuals with different linguistic repertoires must be able to communicate with each other, either in the form of written communication (e.g. meeting documents; *note verbale*) or orally, (for example, in working/correspondence groups or general assembly). In these contexts, different communication strategies are necessary to ensure effective communication; these coexist with the use of a *lingua franca*.

To date some research has addressed the question of the effective representation of linguistic minorities in the public administration of multilingual countries in general (Naff and Jurée Capers 2014; Turgeon and Gagnon 2013; Kübler et al. 2011); whilst others have discussed the influence of cultural and linguistic diversity on public service motivation (Ritz and Brewer 2013) or on public administrators' training (Kolisnichenko and Rosenbaum 2009); other research has addressed the question of how to administrate the electoral vote in multilingual constituencies (Hall 2013). Despite recent interest in the effectiveness/parity of multilingual communication, no studies have attempted to develop indicators which are able to measure the effectiveness of multilingual communication in communities of practice (e.g. work places).

We therefore propose a set of new indicators for measuring diversity in multilingual communication that can be employed in empirical research in multilingual countries and organisations. Such indicators depart from the assumption that effective communication can occur only through a single common language. Our indicators explicitly take into account the possibility of relying on the receptive multilingual skills of speakers and on linguistic mediation services such as interpretation and translation.<sup>2</sup>

<sup>2</sup> On the relationship between translation policy and language policy in general, see Meylaerts and González Nuñez (2017), and Grin (2017).

The remainder of the article is organised as follows. In Sect. 2, we present a group of indicators that compute the probability that people living or working in a multilingual environment, given their linguistic repertoires, the size of the group and the frequency of interaction, can communicate together following different models, i.e.: a common language, receptive multilingualism or a combination of the two. These indices can be used both to measure diversity in a multilingual organisation where people from different linguistic backgrounds work together, as well as multilingual territories where people interact. In Sect. 3, we present two indicators that measure the degree of diversity within a language regime, where a language regime is defined as the language policy of an organisation that determines a set of official and working languages along with rules concerning their use for communication within and outside a multilingual organisation, and the extent of translation and interpreting to be provided in such languages. These indices measure the extent to which documents are translated or oral interventions are interpreted (and therefore available) into the official/working languages of the organisation considered. Section 4 discusses some potential applications of our indices, whilst Sect. 5 concludes our discussion and proposes some lines of enquiry for future investigation.

## 2 Diversity in Multilingual Groups

The original Greenberg's *A* index of linguistic diversity (or monolingual non-weighted method) as Ginsburgh and Weber (2016a) note, was first published by Gini (1912), but Greenberg was the first to apply it to the measurement of linguistic diversity. The simplest computation of balance is the Herfindahl–Hirschman concentration index and is defined as follows:

$$C = \sum_{l=1}^L n_l^2, \quad (1)$$

where  $n_l$  is the population share of group  $l$  (or a firm's market share) and  $L$  is the number of groups (firms) in the population (market) considered. The Simpson diversity index in ecology, the ELF index in economics, and Greenberg's *A* index of linguistic diversity are, in practice, equivalent to  $1 - C$ , where  $L$  is the number of languages and  $n_l$  is the proportion of the population speaking language  $l$  ( $n_l = \frac{N_l}{N}$ , where  $N_l$  is the absolute number of speakers of language  $l$  and  $N$  the total population). For this index all individuals are either considered as monolinguals or only their first language is considered. Greenberg's *A* index is interpreted as the probability that an individual of the population does not share the same language with another randomly selected individual. This population could be the inhabitants of a certain region or all the people working in a certain organisation.

The Greenberg's *A* index, in essence, deals with communication between two monolingual interlocutors. If one is mainly interested in ethno-linguistic diversity, applying the technically easier monolingual indices can be justified, because the underlying assumption is that, although being multilingual, most people belong to only one ethno-linguistic group. This index, nevertheless, is not sufficient to measure diversity in multilingual communication. People living in a multilingual country or working in international organisations are often polyglot. Communication can follow different patterns, including the use of a common language, either a language spoken by a certain percentage of the staff as mother tongue or a lingua franca, or a communication mode in which receptive skills are exploited.

In his 1956 paper, Greenberg already proposed an index—the *H* index—to measure the probability that multilingual people share a common language. Nevertheless, Greenberg’s *H* index is designed to examine the likelihood of successful communication only between two interlocutors. In his extension of the Greenberg indices, Lieberman (1964) investigated communication not between two random members of a society, but between two individuals belonging to different ethnolinguistic groups. For this reason, Lieberman’s indicator is not relevant in the modelling of communication in multilingual organisations.

In order to evaluate the need for language services (i.e. translation and interpreting) as well as language training in multilingual groups, we need indicators that translate data on the linguistic repertoire of people into a synthetic measure of potential effectiveness of communication. In this section, we discuss oral communication among two or more interlocutors speaking different languages. These indices can be easily applied to written communication also. We develop indices for three different models of communication: communication in a shared language, “polyglottism”, and receptive multilingualism.

For the first model (communication in a single language) to be feasible, there must be at least one language in which all group members—or people in the meeting—have sufficient active (productive) and receptive skills; this may be just one single language or more than one. For the index it is irrelevant which single language is chosen for communication in the group. The common language can be the native/preferred language of the majority in the group, the native language of just one group member who does not have sufficient skills in any other language, or an agreed-upon lingua franca.

The second mode, which we call “polyglottism”, enables interlocutors to make use of their linguistic competence in more than one language, including receptive knowledge in the languages spoken by others. Within this model the individual may speak either in one of the shared languages amongst the group or in their native language relying on receptive linguistic competence of colleagues. Hence, this mode of communication takes advantage of different active and receptive skills among the group members. The “polyglottism” mode of communication is essentially a combination of receptive multilingualism (see below) and communication in a single language, and it includes the possibility of code-switching between languages. Other possible modes are neglected here.<sup>3</sup>

The third communication mode is inspired by the “Swiss model”. This is a model of communication in which interactants rely on their receptive language skills when interacting with speakers who employ a different language variety than theirs. Mutual understanding is achieved due to hearer’s/reader’s receptive understanding of the variety/varieties used by their co-interactants, and no common language is required in this model (for literature on receptive multilingualism, see for example, Ten Thije and Zeevaert 2007; Rehbein et al. 2011). This mode of communication is sometimes preferred to the use of a lingua franca as it enables individuals to express themselves in a language of their choice, whilst also accommodating to the preferred language of their interlocutor (assuming that both have sufficient competence in both varieties to facilitate communication). Research on receptive multilingualism is steadily growing, particularly in the fields of bilingualism, contact linguistics, pragmatics, language acquisition and intercultural communication (see Braunmüller 2013; Werlen 2007). Different contexts of use have been explored ranging from macro accounts of multilingual communication in national territories in which

<sup>3</sup> Among these neglected modes are the use of external translators or interpreters, and translation and interpretation by some of the group members. Some indices to measure diversity in multilingual communication when such models are adopted, nevertheless, are discussed in Sect. 3 of this article.

public administrations encourage clients to use their preferred native languages (e.g. see Coray et al. 2015; Christopher Guerra and Zurbruggen 2013), to micro accounts of practices within specific communities of practice in the workplace (Berthele and Wittlin 2013; Mondada et al. 2013; Wodak et al. 2012). Nevertheless, the measurement of success of this communication mode with respect to the “one common language model” has remained almost uncharted so far.

It should be emphasised, however, that for the proposed indicators certain simplifications of the complex reality of multilingual communication are unavoidable. First, we rely on quantitative information on linguistic skills. To use a language in communication—in an active or receptive fashion—*sufficient* skills are needed. But what is sufficient is a question of definition and/or assessment. The definition of sufficient skills should be based on the nature of the communication (simple organisational questions versus in-depth discussion of complex processes), and hence provided by the institution itself and actors therein (e.g. representatives of Governments or Secretariat in the case of the EU or UN). Second, information on language skills within organisations or in multilingual countries is most often obtained via self-assessment (e.g. see the UN’s criteria of assessment for Secretariat personnel, or general surveys on the population), and rarely through language tests. This raises the question of reliability of data and the precise definition of named language. Nonetheless, within the boundaries of quantitative models we conceptualize “multilingual competence as an integrated whole, formed by partial competences in all the varieties (languages and dialects)<sup>4</sup> that the repertoire of the multilingual person consist of [...]” (Lüdi 2007: 173). It is worth noting that we do not study whether people behave as the model prescribes, but only whether communication is possible given a certain distribution of linguistic repertoires. People may follow other communication patterns, especially if they do not know the language repertoire of other interlocutors. Third, we consider three idealised modes of communication recognised within the functioning and language planning mechanisms of organisations themselves. Whilst acknowledging their existence and affordances, we do not—at least not explicitly—account for the complex and dynamic practices of code-switching or translanguaging, an often observed phenomenon in multilingual settings (e.g. Gardner-Chloros 2009; García and Wei 2014). The purpose of this article, nevertheless, is not to explore the possible ways in which situated actors deal with linguistic diversity in professional settings, but rather to provide measurable indices that can be used to compare different contexts and can be applied by decision makers to plan and monitor language policy interventions at the institutional level, in particular with respect to the needs of language services (interpreters/translators), and language training. The complete formulas to compute the indices are provided in Appendix 1. In the next section we study the properties of the indices and we provide some numerical examples.

## 2.1 Communication in a Common Language

As a first index we consider the probability that in a group there is at least one language spoken and understood by all members. We differentiate between two types of cases: individuals either have sufficient competence in a language (active *and* receptive) or they do not. Hence, only having receptive skills is not sufficient to be counted as a speaker of a

<sup>4</sup> We restrict ourselves to named languages only. In principle, the indicators could easily be extended to include different varieties of the same language.

language. We assume that it is not possible to have productive skills without also having receptive skills (we disregard the case of deaf signing individuals who may actually have receptive knowledge of a spoken language but be unable to speak). For every individual  $i$  and language  $l$  we have a variable  $\alpha_l^i$ , where  $\alpha_l^i = 1$  if individual  $i$  has sufficient receptive and productive skills in language  $l$  and  $\alpha_l^i = 0$  otherwise. Then, the linguistic repertoire of individual  $i$  is a vector  $\alpha^i = (\alpha_1^i, \dots, \alpha_L^i)$  comprised of zeros and ones.

First, we calculate the probability that in a randomly composed group of  $m$  people there is a common language. We denote this probability by  $P_{com}^m$ . This is very similar to Greenberg’s  $H$  index and the probability  $W_F^N(M)$  used in Voslamber (2018).<sup>5</sup> Recall, Greenberg considers the same probability, but only for  $m = 2$ . In Voslamber (2018),  $W_F^N(M)$  is the probability that in a meeting of  $M$  people ( $M \geq 2$ ) there is a common language if  $N$  working languages are assigned, of which each staff member has to know  $F$  (foreign) languages that differ from her first language,  $F < N$ . For our indicators, people can speak differing numbers of languages and individuals can even be monolingual. For an analysis of the so-called “mother tongue + 2” model for language education in the European Union—according to which every EU citizen should learn two foreign languages in addition to his or her mother tongue—Grin (2006) derives the probability of a common language for arbitrary group sizes for the case of three languages (i.e. German, French and English) if language skills are distributed equally (one third knows German and English, one third German and French and one third English and French). Our model has no restrictions regarding the distribution of language skills of the speakers.

The derivation of the formula for  $P_{com}^m$  is provided in the Appendix 1. To derive this probability, one needs information on the linguistic repertoires of people for example, members of staff in an organisation. Next, we can follow two approaches. For the first approach, an estimation of the median meeting size  $\bar{m}$  is needed. Then,  $P_{com}^{\bar{m}}$  is the probability that in a randomly composed meeting of median size  $\bar{m}$  there is a language spoken by all the staff members in the meeting. This yields our first index

$$\phi_{com}^{\bar{m}} = P_{com}^{\bar{m}} \tag{2}$$

For the second approach, an estimation of frequencies of different group composition or meeting sizes and the duration of meetings is needed. For meeting sizes  $m$ , we denote by  $F_m$  the average daily number of meetings of size  $m$  multiplied by the average duration of a meeting of size  $m$ . Moreover, we introduce the fraction  $f_m = F_m / \sum F_i$ , which is a measure of the importance of meetings of size  $m$ . For example, if every day there are on average 30 meetings with two people that last 1 hour each ( $F_2=30$ ), and 25 meetings with three people of 2 hours ( $F_3=50$ ) and ten meetings with four people also of 2 hours ( $F_4=20$ ), then,  $f_2=0.3$  and  $f_3=0.5$  and  $f_4=0.2$ . If the average duration of a meeting is independent of the meeting size, then the fractions  $f_m$  are just the distribution of meeting sizes. For an adjusted version of the first index we weight the probabilities of successful communication by these fractions:

$$\phi_{com} = \sum_m f_m \cdot P_{com}^m \tag{3}$$

Let us provide a simple example. We consider the case of two languages and meetings of just two individuals. Applying the general formula presented in the Appendix 1, we obtain

<sup>5</sup> Voslamber (2018) analyses the effect of different numbers of working languages on failed communication in multilingual teams of different sizes within EU institutions. In addition, he investigates how the probability of failed communication can be reduced if EU staff members know three foreign languages instead of just two.



$P_{com}^2 = 1 - 2n_1n_2$ , where  $n_1$  is the fraction of monolinguals in language 1 and  $n_2$  is the fraction of monolinguals in language 2. Hence, if 63% of the staff is monolingual in language 1, 18% is monolingual in language 2 and 19% is bilingual, then we obtain  $P_{com}^2 = 0.77$ . In contrast, if 63% are monolingual in language 1 while all the others are bilingual, we obtain  $P_{com}^2 = 1$ . As one would expect, in the latter case there are no communication problems, since everybody can communicate with everybody else. As a second example, we consider three languages and meetings of two individuals. We have the following repertoires:

- $R_1 = (1,0,0)$ : competence only in language 1;
- $R_2 = (0,1,0)$ : competence only in language 2;
- $R_3 = (0,0,1)$ : competence only in language 3;
- $R_4 = (1,1,0)$ : competence in languages 1 and 2;
- $R_5 = (1,0,1)$ : competence in languages 1 and 3;
- $R_6 = (0,1,1)$ : competence in languages 2 and 3;
- $R_7 = (1,1,1)$ : competence in all three languages.

Let  $n_j$  be the fraction of people with repertoire  $R_j$ ,  $j = 1, \dots, 7$ . Then, we get

$$P_{com}^2 = 1 - n_1(n_2 + n_3 + 2n_6) - n_2(n_1 + n_3 + 2n_5) - n_3(n_1 + n_2 + 2n_4) \tag{4}$$

The probabilities  $P_{com}^3$  and  $P_{com}^4$  are comparable polynomials of degree three and four, but too lengthy to be presented here. As a numerical example, let  $n_1 = 0.63$ ,  $n_2 = 0.10$ ,  $n_3 = 0$ ,  $n_4 = 0.12$ ,  $n_5 = 0.11$ ,  $n_6 = 0$  and  $n_7 = 0.04$ . For this distribution of skills, we obtain  $P_{com}^2 = 0.85$ ,  $P_{com}^3 = 0.74$  and  $P_{com}^4 = 0.66$ . If we assume the above distribution and duration of meeting sizes ( $f_2 = 0.3$ ,  $f_3 = 0.5$ ,  $f_4 = 0.2$ ), then we get  $\phi_{com}^{\bar{m}} = 0.74$  and  $\phi_{com} = 0.76$ .

## 2.2 Polyglottal Communication

The second index measures the probability that all members of a group can communicate with each other taking advantage of all active and receptive skills within the group. For every individual  $i$  and language  $l$  we have a variable  $\beta_l^i$ , where  $\beta_l^i = 2$  if individual  $i$  has sufficient receptive and productive skills in language  $l$ ,  $\beta_l^i = 1$  if individual  $i$  has only receptive skills in  $l$  and  $\beta_l^i = 0$  else. Here, the linguistic repertoire of individual  $i$  is a vector  $\beta^i = (\beta_1^i, \dots, \beta_L^i)$  comprised of zeros, ones and twos. Note, that the vector  $\beta$  contains more information than the  $\alpha$ . Given  $\beta_l^i$ , we can derive  $\alpha_l^i$  via  $\alpha_l^i = 1$ , if  $\beta_l^i = 2$ , and  $\alpha_l^i = 0$  else. As for the common language, we can derive the probability that in a randomly composed group with  $m$  members, every individual can use a language in which all the other individuals in the group have at least receptive knowledge. This probability of successful communication is denoted by  $P_{poly}^m$ . How to derive  $P_{poly}^m$  is explained in the Appendix 1. As before, based on estimates on the median meeting size  $\bar{m}$  and/or the frequencies of certain groups compositions or meeting size sizes  $p_m$ , we get two indices:

$$\phi_{poly}^{\bar{m}} = P_{poly}^{\bar{m}} \tag{5}$$

If  $m$  can vary, we get:

$$\phi_{poly} = \sum_m f_m \cdot P_{poly}^m \tag{6}$$

As an example, we consider two languages and groups of two people. We have the repertoires

- $R_1 = (2,0)$ , productive skills in language 1, no skills in language 2;
- $R_2 = (0,2)$ , productive skills in language 2, no skills in language 1;
- $R_3 = (2,1)$ , productive skills in language 1, receptive skills in language 2;
- $R_4 = (1,2)$ , productive skills in language 2, receptive skills in language 1;
- $R_5 = (2,2)$ , productive skills in both languages.

The probability that two people who meet randomly can communicate with each other is given by

$$P_{poly}^2 = 1 - 2(n_1n_2 + n_1n_4 + n_2n_3) \tag{7}$$

where  $n_j, j = 1, \dots, 5$ , is the fraction of people with repertoire  $R_j$ . A numerical example for  $m = 2, m = 3$  and  $m = 4$  is provided in Sect. 2.5.

### 2.3 Receptive Multilingualism

For a third index we have devised a way to calculate the probability that individuals in a group can use their first language (presumably the language they prefer to converse in), relying on the receptive skills of other interlocutors (i.e. as in the “Swiss model” of communication). We include measures of active and receptive skills. For every individual  $i$  and language  $l$  we have, as for polyglottal communication, a variable  $\beta_l^i \in \{0, 1, 2\}$ . Furthermore,  $\gamma^i$  contains the information on the preferred or native language of individual  $i$ . Here, an individual  $i$  is characterized by a vector  $\beta^i = (\beta_1^i, \dots, \beta_L^i)$ , comprised of zeros, ones and twos, and a number  $\gamma^i \in \{1, \dots, N\}$ . Based on this information for all the individuals or a representative sample, we can derive the probability that in a group of size  $m$  everybody can use his/her preferred language. This probability is called  $P_{rec}^m$ . In Grin et al. (2015), similar probabilities for the analysis of the functioning of the “Swiss model” are presented, but they are restricted to the probabilities of successful communication if two or three individuals with different first languages meet.<sup>6</sup> The index presented in this article is more general and it does not put any restriction on the number of languages spoken by actors and the number of people involved in a meeting. Based on  $P_{rec}^m$ , we again obtain two indices of successful communication:

$$\phi_{rec}^{\bar{m}} = P_{rec}^{\bar{m}} \tag{8}$$

If  $m$  can vary, we get:

$$\phi_{rec} = \sum_m f_m \cdot P_{rec}^m \tag{9}$$

As an example, we again consider the case of two languages and meetings with two people. We have the following repertoires:

- $R_1 = (1|2, 0)$ , the preferred language is 1, productive skills in 1, no skills in 2;

<sup>6</sup> The authors calculate the probabilities of successful communication if two or three Swiss people with different first languages meet, e.g. a German native (or “L1”) speaker and a French L1 speaker, a German L1 speaker and an Italian L1 speaker, a French L1 speaker and an Italian L1 speaker, or a German L1 speaker, a French L1 speaker and an Italian L1 speaker. Note that this model does not include the possibility that people be bilingual.

- $R_2 = (2|0, 2)$ , the preferred language is 2, productive skills in 2, no skills in 1;
- $R_3 = (1|2, 1)$ , the preferred language is 1, productive skills in 1, receptive skills in 2;
- $R_4 = (2|1, 2)$ , the preferred language is 2, productive skills in 2, receptive skills in 1;
- $R_5 = (1|2, 2)$ , the preferred language is 1, productive skills in both languages;
- $R_6 = (2|2, 2)$ , the preferred language is 2, productive skills in both languages;

Let  $n_j$  be the fraction of people with repertoire  $R_j$ ,  $j = 1, \dots, 6$ . We get,

$$P_{rec}^2 = 1 - n_1(n_2 + 2n_4 + 2n_6) - n_2(n_1 + 2n_3 + 2n_5) \quad (10)$$

A numerical example for  $m = 2$ ,  $m = 3$  and  $m = 4$  is provided in Sect. 2.5.

## 2.4 Properties of the Potentially Successful Communication Indices

The indices  $\phi_{com}^{\bar{m}}$ ,  $\phi_{rec}^{\bar{m}}$ ,  $\phi_{poly}^{\bar{m}}$  and  $\phi_{com}$ ,  $\phi_{rec}$ ,  $\phi_{poly}$  all satisfy the following properties:

1. The index is a number between 0 and 1.
2. The higher the probability for successful communication (either defined as common language or use of preferred language), the higher the index.
3. The higher the median group size  $\bar{m}$ , the lower the median index  $\phi^{\bar{m}}$ .
4. Since all language skills can be exploited to guarantee successful communication, the index ( $\phi_{poly}$ ) is always the highest of the three. This is an important point: the Swiss model needs more support than polyglottism to be effective. Nevertheless, if people have sufficient receptive skills, then the Swiss model can theoretically work with a high number of languages. Whether  $\phi_{com}$  or  $\phi_{rec}$  is higher depends on the distribution of active and receptive language skills, (see the numerical example in Sect. 2.5). If a high percentage of people have receptive skills in the majority of the working languages of the organisation or the official languages of a country, then  $\phi_{rec}$  tends to be the highest of the two indices.

Indices assist policy makers by providing guidance about choices. Calculated for all (administrative) units (e.g. different departments) of an organisation or districts of a territory, the index can be used to identify those units for which intervention is needed most (i.e. those with the lowest index numbers). It is worth noting that the three indicators measure the probability of successful communication in their respective mode of communication, but they do not point out the distributive consequences of alternative ways of handling multilingual communication. If a single common language is used, for example, it can be the first language of one interlocutor, but the second language of all the others. This can happen also in the polyglottal mode. Alternatively, in the receptive mode, communication can be effective by allowing speakers to use their preferred language. This illustrates that successful communication in one of these two modes does not imply the same level of equity among the different interlocutors. This question is very much relevant in the evaluation of language regimes (Gazzola 2014), but it is not addressed in this article.

## 2.5 Illustration

To illustrate the three indices, we now apply them to a numerical example (see Table 1). Consider two languages  $A$  and  $B$ . We assume that 70% of the population have  $A$  as their first/preferred language and that the remaining 30% have  $B$  as their first language. Of those

**Table 1** Example of distribution of language skills in a hypothetical population

Monolingual		With receptive skills		Bilingual	
<i>A</i> (%)	<i>B</i> (%)	<i>A, B<sub>r</sub></i> (%)	<i>B, A<sub>r</sub></i> (%)	<i>AB</i> (%)	<i>BA</i> (%)
35	3	28	15	7	12

**Table 2** Indicators of potentially successful communication for the three different communication models (figures rounded at the second decimal)

Communication model	$m=2$	$m = \bar{m} = 3$	$m=4$	Weighted by $f_m^*$
Common language ( $\phi_{com}^m$ )	0.77	0.60	0.47	0.62
Polyglottism ( $\phi_{poly}^m$ )	0.86	0.70	0.56	0.72
Receptive multilingualism ( $\phi_{rec}^m$ )	0.77	0.55	0.38	0.58

\*This is a weighted indicator that takes the different sizes of the groups into account. The percentages were obtained from an implementation of formulas (3), (6) and (9) in the software MatLab. The code is available on request directly from the authors

having *A* as their first language, 50% are fully monolingual, 40% have only receptive skills in language *B* (*B<sub>r</sub>*) and 10% are fully bilingual *AB* (having productive and receptive skills in both languages). Of those having *B* as their first language, 10% are monolingual, 50% have receptive skills in *A* (*A<sub>r</sub>*) and the remaining 40% are bilingual (*BA*). With regard to the entire population, the distribution of language skills is reported in Table 1.

We assume that 30% of all communication situations involve two people, 50% involve three people and 20% involve four people, and that the average duration of meetings is independent of the meeting size. Therefore,  $f_2 = 0.3$ ,  $f_3 = 0.5$  and  $f_4 = 0.2$ . Consequently, the median group  $\bar{m}$  size is three.

In Table 2 all three indices are listed in relation to the different group sizes as well as for the weighted case. As one would expect, the higher the size, the lower is the probability of successful communication in all three modes. Moreover,  $\phi_{poly}$  is always higher than the other two indices. This happens because communication in a common language and communication in everyone's preferred language are more restrictive modes of communication. We can see that making use of the entire linguistic repertoire instead of just one language increases the probability of successful communication by 10%. That  $\phi_{rec}$  is lower than  $\phi_{com}$  is an effect of the special distribution of repertoires considered here. If a larger number of *A* speaking people had receptive skills in *B*, then it would be the other way around. Due to the special distribution of group sizes assumed here, the median indicators are slightly lower than the weighted indicators. If only 20% of all groups involve two interlocutors and 30% involve three, then the opposite would be true.

### 3 Diversity in Multilingual Language Regimes

Greenberg's *A* index and similar indices such as the Simpson index or the Shannon entropy index—an indicator used in ecology that combines species richness and their relative abundance—measure diversity in a given environment under the assumption that observable units belong only to one group (e.g. species or languages). This assumption is not realistic in contexts where observable units can belong at the same time to many groups. While

it is unlikely that all individuals within a population speak all languages (if the number of languages is relatively high), it is perfectly possible that all documents are translated into all the official languages of a state or an international organisation. Hence, considering potential interactions between pairs of individuals is not useful to measure the degree of diversity of translation policies of and in multilingual organisations, because the unit of observation in this case is not the individual, but documents through which multilingual communication happens.<sup>7</sup> It is necessary, therefore, to develop indicators that can help decision makers to compare the degree of documents' linguistic diversity (as opposed to diversity in an ecosystem) in contexts where translation can be provided into all official languages.

In our view, two criteria should be combined. First, all other things being equal and given a set of official languages, a regime  $l$  is more linguistically diverse than regime 2 if the proportion of documents translated in  $l$  is higher than in 2. For a first index, we assume that documents are produced by default in one language (the “default language”) and are translated into  $L$  other official or working languages (this assumption is relaxed later).  $D$  denotes the total amount of documents produced in the default language, while  $D_l$  is the number of documents translated into language  $l$ . Hence,  $d_l = \frac{D_l}{D}$  is the percentage of documents translated into language  $l$ . The first criterion can be operationalised through a simple indicator that denotes the average percentage of translated documents. We call this indicator “average” ( $\mu$ ), and it is computed as follows:

$$\mu = \frac{1}{L} \sum_{l=1}^L d_l \quad (11)$$

We assume that  $d_l$  is strictly positive ( $d_l > 0$ ), because it would be nonsense to declare a language as official if it is never used in practice. This means that  $0 < \mu \leq 1$ .

The second criterion embodies the variance of the distribution of documents translated into different languages. Assume that regime 1 translates 99% of documents into language  $A$  and only 1% into language  $B$ , whereas language regime 2 translates 50% of documents into both languages. The average ( $\mu$ ) is the same in both cases, but it would be misleading to claim that they are equally multilingual, as in regime 1 language  $B$  is barely used. In order to take this into account, therefore, we need an indicator that gives a higher ranking to language regime 2 than to language regime 1, all other things being equal. We define the “polarisation index” ( $\rho$ ) as:

$$\rho = 1 - \frac{1}{L} \sum_{l=1}^L (1 - d_l)^2 \quad (12)$$

where  $0 < \rho \leq 1$ . The indicator  $\rho$  measures the average squared deviation from a full translation regime. In other words,  $d_l$  is the percentage of documents actually translated into language  $l$ , and  $1 - d_l$  the “distance” from full translation. If all documents are translated into all official languages the value of  $\rho$  is 1, i.e., all languages are treated as the same level as the default language. Hence, the larger the value of  $\rho$ , the more a language regime approaches full translation and, therefore, the lower polarisation. This indicator is not a simple indicator of equality or variance with respect to the mean. A simple indicator of

<sup>7</sup> Gazzola (2014: 138–139) discusses the pitfalls of using the Herfindahl–Hirschman concentration index as a single measure of diversity in multilingual regimes.

**Table 3** Measuring multilingualism in language regimes, examples

Language regime	A	B	C	D	E	F
Language (excluding the default language)	Proportion of documents translated					
1	100%	50%	100%	100%	33%	100%
2	100%	50%	75%	100%	33%	75%
3	100%	50%	50%	48%	33%	10%
4	100%	50%	24%	1%	33%	10%
5	100%	50%	1%	1%	33%	10%
$\mu$	1.00	0.50	0.50	0.50	0.33	0.41
$\rho$	1.00	0.75	0.63	0.55	0.55	0.50

variance, in fact, is not able to capture the difference between a language regime in which 90% of documents are equally translated into all the working languages of an organisation, and a language regime in which only 10% of official documents are equally translated into all working languages. By contrast, the polarisation index captures such differences: the larger the gap among languages in terms of the difference between full and actual translation, the lower the index. We can summarize the properties of  $\mu$  and  $\rho$  as follows:

- Both take a value between 0 and 1, and the value 1 is obtained when full translation into all official languages is provided.
- The polarisation index  $\rho$  is a positive function of the mean  $\mu$  and it is negatively correlated with concentration, measured by the Herfindahl–Hirschman concentration index  $C$  (see Appendix 2)

If the language regime translates only a limited number of documents  $D_{max}$  from the default language into the other languages (so  $d_l < 1$ ), i.e.  $D_1 + \dots + D_L = D_{max}$ , then the polarization index  $\rho$  is maximal if the translation effort is evenly distributed over all the languages. That is  $D_1 = \dots = D_L = D_{max}/L$  (see Appendix 2).

The combined use of  $\mu$  and  $\rho$  can be used to compare regimes, to rank-order them, and to clarify trade-offs. In some cases the application of the indices lead to indeterminate outcomes. We can therefore derive the following rules: a regime  $X$  is more linguistically diverse than regime  $Y$  if  $\mu_X > \mu_Y$  and  $\rho_X \geq \rho_Y$ ; or if  $\mu_X \geq \mu_Y$  and  $\rho_X > \rho_Y$ . By contrast, if  $\mu_X > \mu_Y$  and  $\rho_X < \rho_Y$ , or if  $\mu_X < \mu_Y$  and  $\rho_X > \rho_Y$  no conclusive results can be obtained, and decision makers must weigh trade-offs.

Table 3 presents an example of seven hypothetical language regimes, and the corresponding values of  $\rho$  and  $\mu$ .

Clearly, regime  $A$  (full multilingualism) is more linguistically diverse than all other regimes. The value of  $\mu$  in regime  $B$ ,  $C$  and  $D$  is the same.  $D$  is more polarised than  $C$  (that is,  $\rho_D < \rho_C$ ), and  $C$  is more polarised than  $B$  (that is,  $\rho_C < \rho_B$ ). This is due to the fact that in regime  $C$  there is just one language into which almost no documents are translated (language 5), a language into which all documents are translated (1) and three languages into which translation is provided (often or quite often). In language regime  $D$  a marginal percentage of documents is available in languages 4 and 5, and there are two languages (1 and 2) into which all documents are translated. Given that  $\mu$  is the same for regimes  $B$ ,  $C$ , and  $D$ , the three regimes can be rank-ordered according to the value of  $\rho$ . As a result,  $B$  is more linguistically diverse than  $C$  and  $C$  more diverse than  $D$ . Regime

$D$  is as polarised as  $E$ , but the value of  $\mu$  is higher in  $D$  than in  $E$ . As a result, the former is more linguistically diverse than the latter. Regime  $F$  translates on average a higher proportion of documents than  $E$ , but it is more polarised (that is,  $\mu_F > \mu_E$  and  $\rho_F < \rho_E$ ). Hence, they cannot be rank-ordered only applying the two criteria.

We relax now the assumption that there is one (and only one) default language. In many multilingual organisations source documents resulting from deliberation are available in different languages and then these documents may (or may not) be translated. Let us define  $D_l^s$  as the number of all documents that were originally drafted in language  $l$  (the superscript  $s$  stands for “source language”) and  $D_l^t$  as the number of documents that are translated into language  $l$  from other source languages (the superscript  $t$  stands for “target language”). The variable  $D_l$  is now the number of documents available in language  $l$  ( $D_l = D_l^s + D_l^t$ ), and the variable  $L$  denotes the total number of working languages (and not only the language into which translation is provided from a default language).  $D^*$  denotes the total amount of original draft documents produced. Hence  $d_l^* = \frac{D_l}{D^*}$  is the percentage of documents available in language  $l$ .  $\mu^*$  is computed as follows:

$$\mu^* = \frac{1}{L} \sum_{l=1}^L d_l^* \quad (13)$$

If there is no default language all documents are first written in, the “polarisation index” ( $\rho^*$ ) is defined as follows:

$$\rho^* = 1 - \frac{1}{L} \sum_{l=1}^L (1 - d_l^*)^2 \quad (14)$$

where  $1 - d_l^*$  denotes the difference between 100% and the actual percentage of documents available in language  $l$ . 100% represents the (theoretical) maximum achievable.

The indices presented in this section can be used also to measure linguistic diversity of interpretation regimes. In this case,  $D_l^s$  is the number of any oral interventions made in language  $l$  and  $D_l^t$  the number of interventions interpreted into language  $l$ . From the point of view of an  $l$ -speaker what matters is the percentage of interventions he/she can hear in language  $l$ .

Two final remarks are in order. First, our indices do not take variety into account, that is, they are not meant to compare the degree of multilingualism of regimes that do not have the same number of official languages. Second, we do not consider external communication, that is, the effects of translation policy on access to official documents by external agents (e.g. citizens or companies).<sup>8</sup>

## 4 Potential Applications

The need for multilingualism management indices is pressing. First, these indices are highly relevant to the study of the political and economic implications of linguistic diversity in different countries. Research in economics and political science, as shown in the

<sup>8</sup> For indicators to measure the degree of linguistic exclusion (or linguistic disenfranchisement) in the population due to language policy, please see, Ginsburgh and Weber (2005) and Gazzola (2016a).

introduction of this article, tends to use fractionalisation indices such as the Greenberg's *A* index (or the *B* index to take disparity into account) as a proxy for linguistic and ethnic fragmentation to explore the impact of linguistic diversity on political, economic and social variables. Greenberg argued that "our general expectation is that areas of high linguistic diversity will be those in which communication is poor, and that the increase of communication that goes with greater economic productivity and more extensive political organisation will typically lead to the ultimate disappearance of all except a single language" (1956: 110). Most papers published on this topic point out that linguistic fragmentation has indeed a negative impact on economic development or social cohesion.<sup>9</sup> Linguistic diversity, nevertheless, can be managed through language policy. People can learn new languages, they can be encouraged to use their receptive as well as productive language skills and repertoires in order to better accommodate people speaking other languages, therefore reducing misunderstandings and potential sources of conflict. Public authorities can provide official documents, collective goods such as road signs or broadcasting, and public services such as health care in more than one language. The proposed indices could be used to evaluate the assumption that it is not language diversity per se that has a negative impact on economic development or political unity; it is the way in which linguistic diversity is managed that makes the difference (on this topic see Liu and Pizzi 2018). The indicators presented in this study, in fact, provide a means of measuring the probability of successful communication instead of the simplistic index of ethnolinguistic fractionalisation which has been employed to date. Communication can be effective and smooth even if linguistic diversity is high. This emphasises the importance of language policy and planning, and therefore the role of the state/public administration in managing linguistic diversity in effective ways. The indicators identify new explanatory variables to study the impact of language diversity on political, administrative and economic outcomes.

Second, indicators can be applied in the study of linguistic diversity management in multilingual organisations and public administrations. Multilingualism is a central policy dimension in public administration of multilingual countries such as South Africa, India, Switzerland and Canada, or multilingual regions such as Wales or Catalonia. The legislation in Switzerland and Canada, for example, requires that languages should be treated on an equal footing in the federal public administration.<sup>10</sup> Empirical research and official reports, nevertheless, have shown that the relationship between the official languages (and therefore their speakers) is characterised by substantial inequality at different levels, including the use of languages in meetings, the level of competence of civil servants in the second and third languages, the representation of linguistic communities in senior positions.<sup>11</sup> Surprisingly, no indicator has been developed to quantify the likelihood that communication in more than one language can work in practice. Without this piece of information, nevertheless, it is not possible to correctly assess the need for language policy and training in the units of the federal administration.

<sup>9</sup> In addition to references quoted in the introduction, see Ginsburgh and Weber (2016a), Arcand and Grin (2013), and Nettle (2000) for a general discussion.

<sup>10</sup> See the Swiss *Federal Act on the National Languages and Understanding between the Linguistic Communities* adopted in 2007 and the related *Ordinance* (2010), and the Canadian *Official Languages Act*, in force since 1969, and modified through time (1988 and 2005).

<sup>11</sup> For Switzerland see Kübler and Zwicky (2018), Gazzola (2016b), DFP (2015) and its annexes, Kübler et al. (2011), Coray et al. (2015) and Christopher Guerra and Zurbriggen (2013). For Canada, see Borbey and Mendelsohn (2017), Cardinal (2015), Gaspard (2015) and Turgeon and Gagnon (2015).



At the international level, the report drafted by Ehmke-Gendron (2015) for the translation service of the Council of the European Union and the related critical internal note published by the “Groupe T2020” (2016) addresses the issue of the measurement of linguistic diversity in translation policy, and identifies the need for measuring the degree of multilingualism of the set of documents they produce and publish. This equally applies to the functioning of the United Nations. A recent review of the way in which multilingualism is managed across the UN system (McEntee-Atalianis 2015; General Assembly 2017) has revealed the need to reform the organisation’s language policy. Global, pragmatic, political and recent economic constraints have led to ever greater lingua franca usage (particularly English) within the organisation, despite calls by the organisation’s secretariat and member states to counter the ecological imbalance amongst the working and official languages and the increasing hegemony of English (Kudryavtsev and Ouedraogo 2003; Fall and Zhang 2011). For changes to be made to the current systems and for principled analyses of current working practices to be undertaken, detailed mathematical modelling of (alternative) language regimes to support bespoke organisational needs for meetings is needed, i.e. analyses of communication across and within different levels and layers of the organisation, such as plenary meetings; working and correspondence groups; and field work activities.

In the European Union there is no formal distinction between official and working languages (Van der Jeught 2015), and therefore any of the 24 official languages can be used in internal meetings in some of its institutions. A restricted number may be used in preparatory meetings, working parties or for internal operations. Limits are imposed according to budgetary and practical constraints. Clearly, communication can be difficult or even impossible if civil servants or people who temporarily work in a multilingual organisation (e.g. the Member of the European Parliament and their assistants) do not share a common language or do not have adequate receptive competences in the language of their colleagues (see, for example, Podestà 2001 for a discussion of the linguistic challenges of the enlargement of the EU with the inclusion of 10 new Member States in 2004, and Kruse and Ammon 2018).

## 5 Conclusions and Directions for Future Research

This article has presented new indices to measure diversity in multilingual communication. These indices offer a way to measure the degree of diversity of communication based on the affordances of translation and interpretation (adopted by multilingual organisations), as well as the means to measure the probability that people can effectively communicate either via one common language or by relying on their receptive competence in more than one language. We acknowledge that the sociolinguistic situation on the ground is more complex as it includes issues such as code-switching and translanguaging. Actors do not always meet by chance but because they are part of a network in which they share interests and goals. The information about the language skills of the other interlocutors may be incomplete, and path-dependences play a role in explaining patterns of language use. However, the indices presented here do improve the measurement of such multilingual communicative contexts by capturing significant variables that have been previously overlooked. Moreover disparity (or distance) between languages could be taken into account by our indices, weighting them by coefficients that reflect the degree of similarity between languages.

The proposed indicators represent valuable tools for the assessment of communication barriers and problems in multilingual regions and organisations when actors (e.g. citizens or civil servants) can be determined, and where data allow computations to determine the extent to which communication in a given mode is possible. If, for example, second language competence of personnel is insufficient, then the “Swiss model” is not applicable and policy intervention might be needed to support speakers of a minority language to use their language at work. Our indices can be combined in order to better inform language policy. For example, indicators of the probability of successful communication in multilingual meetings discussed in Sect. 2 can be employed to plan the provision of interpreting services in an organisation, and the indicators of diversity of language regimes presented in Sect. 3 can be used to monitor the implementation of such plan. Hence, the indices make a valuable contribution to language policy design, implementation and evaluation supporting recent calls for evaluative frameworks.

**Acknowledgements** The financial support of the European Union’s Seventh Framework Program (Project MIME—Grant Agreement 613344) and of Birkbeck, University of London is gratefully acknowledged. The authors wish to thank Bengt-Arne Wickström, Dietrich Voslamber and two anonymous reviewers for their valuable suggestions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix 1: Derivation of Probabilities $P_{com}^m$ , $P_{poly}^m$ , and $P_{rec}^m$

In the following, we provide formulas for the probabilities  $P_{com}^m$ ,  $P_{poly}^m$ , and  $P_{rec}^m$ . Recall, we consider  $L$  languages and a meeting size  $m$ . In the following, for simplicity we use the superscripts  $c, p, r$  instead of the subscripts  $com, poly, rec$ .

#### Common Language: $P_{com}^m$

First, we have to determine the number of possible language repertoires. Recall, in this case an individual either speaks a language or does not—we do not differentiate between active and receptive skills. Each individual can be monolingual or speak up to  $L$  languages. The number of possible language repertoires is thus given by:

$$r^c(L) = \sum_{k=1}^L \binom{L}{k} \tag{15}$$

For  $L = 2, 3, 4$  we obtain  $r^c(2) = 3$ ,  $r^c(3) = 7$  and  $r^c(4) = 15$ . As the  $\alpha^i$ , a language repertoire is a  $L$ -dimensional vector consisting of zeros and ones. We number the language repertoire types and call them  $R_1^c, \dots, R_{r^c(L)}^c \in \{0, 1\}^L$ . For every language repertoire type  $R_s^c$  we now define the set of language repertoire types that have at least one language in common with  $R_s^c$ :  $\Omega^c(s) = \{R_i^c | R_i^c \cdot R_s^c \neq 0\}$ , where the sign “ $\cdot$ ” denotes the scalar product. Obviously,  $R_s^c \in \Omega^c(s)$ . Moreover, for  $s_1, \dots, s_k$  we define the intersection  $\Omega^c(s_1, \dots, s_k) = \bigcap_{j=1}^k \Omega^c(s_j)$  and the set of all language repertoire types  $\Omega^c = R_1^c, \dots, R_{r^c(L)}^c$ .

Next, we derive the distribution of language repertoire types within the institution of interest. Let  $N_s$  denote the number of individuals with language repertoire  $R_s^c$ . By  $n_s = N_s/N$  we denote the fraction of individuals with language repertoire  $R_s^c$ . Given this information, we can calculate the probability that  $m$  people who meet randomly (e.g. in a meeting) speak a common language:

$$P_{com}^m = \sum_{R_{s_1}^c \in \Omega^c} n_{s_1} \sum_{R_{s_2}^c \in \Omega^c(s_1)} n_{s_2} \sum_{R_{s_3}^c \in \Omega^c(s_1, s_2)} n_{s_3} \cdots \sum_{R_{s_m}^c \in \Omega^c(s_1, s_2, \dots, s_{m-1})} n_{s_m} \tag{16}$$

Note that  $\sum_{R_{s_1}^c \in \Omega^c} n_{s_1} = 1$ .

**Polyglottal Communication:  $P_{poly}^m$**

As in the case of the common language index, we have to define all possible language repertoire types. Since receptive and active knowledge are considered here, there are more possible language repertoire types:

$$r^p(L) = \sum_{k=1}^L \binom{L}{k} \sum_{l=0}^{L-k} \binom{L-k}{l} \tag{17}$$

For  $L = 2, 3, 4$  we obtain  $r^p(2) = 5$ ,  $r^p(3) = 19$  and  $r^p(4) = 65$ . As the  $\beta^i$ , a language repertoire is a  $L$ -dimensional vector consisting of zeros, ones and twos. We number the language repertoire types and call them  $R_1^p, \dots, R_{r^p(L)}^p \in \{0, 1, 2\}^L$ . Moreover, for every language repertoire type  $R_s^p$  we define a corresponding vector  $\bar{R}_s^p \in \{0, 2\}^L$  as follows: the  $j$ 'th component of  $\bar{R}_s^p$  is 2, if the  $j$ 'th component of  $R_s^p$  is 2 (productive and receptive skills), and 0 else. For every language repertoire type  $R_s^p$  we now define the set language repertoire types have at least receptive skills in one of the active languages of repertoire type  $R_s^p$ :

$$\Theta^p(s) = \{R_u^p | \bar{R}_s^p \cdot R_u^p > 0\} \tag{18}$$

Since not only the individual of type  $R_s^p$  should be able to use one of the languages she has productive skills in, but also the individuals she is communicating with, we define  $\Omega^p(s) = \{R_u^p | R_u^p \in \Theta^p(s) \wedge R_s^p \in \Theta^p(u)\}$ . Note, repertoires  $R_u^p \in \Omega^p(s)$  if  $R_u^p$  has at least receptive skills in one of  $R_s^p$ 's active languages and if  $R_s^p$  has at least receptive skills in one of  $R_u^p$ 's active languages. That is, both can communicate with each other employing their active and receptive skills. As before, we also define  $\Omega^p(s_1, \dots, s_k) = \Omega^p(s_1) \cap \dots \cap \Omega^p(s_k)$  and  $\Omega^p = \{R_1^p, \dots, R_{r^p(L)}^p\}$ .

Let  $N_s$  be the number of individuals with language repertoire type  $R_s^p$ . The fraction of individuals being of type  $R_s^p$  equals  $n_s = N_s/N$ . The probability that  $m$  individuals who meet randomly can all use their preferred languages is then given by:

$$P_{poly}^m = \sum_{R_{s_1}^p \in \Omega^p} n_{s_1} \sum_{R_{s_2}^p \in \Omega^p(s_1)} n_{s_2} \sum_{R_{s_3}^p \in \Omega^p(s_1, s_2)} n_{s_3} \cdots \sum_{R_{s_m}^p \in \Omega^p(s_1, s_2, \dots, s_{m-1})} n_{s_m} \tag{19}$$

**Receptive Multilingualism:  $P_{rec}^m$**

In the third case, for every individual  $i$  we have two pieces of information: their language skills and their preferred language. We have a linguistic repertoire  $\beta^i = (\beta_1, \dots, \beta_L)$ . The number of language repertoire types is the same as for  $P_{poly}^m$ . Moreover, by  $\gamma^i \in \{1, \dots, L\}$  we denote the

preferred language of individual  $i$ . It is reasonable to assume that the individual has productive skills in her preferred language, i.e.  $\beta_{y^i}^i = 2$ . Then, the number of individual types, that is language repertoire plus preferred language, equals

$$t(L) = \sum_{k=1}^L k \binom{L}{k} \sum_{l=0}^{L-k} \binom{L-k}{l} \tag{20}$$

For  $L = 2, 3, 4$  we obtain  $t(2) = 6$ ,  $t(3) = 27$  and  $t(4) = 108$ . Every individual type corresponds to a vector  $T_s = (T_s^0, T_s^1, \dots, T_s^L) \in \{1, \dots, L\} \times \{0, 1, 2\}^L$ , where  $T_s^0$  is the preferred language of individual type  $T_s$ . For every such type  $T_s$  we define in the set of all types that are compatible with  $T_s$ , that is all types that have at least receptive knowledge of type  $T_s$ 's preferred language:

$$\Theta^r(s) = \{T_u | T_u^{T_s^0} > 0\} \tag{21}$$

Since not only the individual of type  $T_s$  should be able to use her preferred language, but also the individuals he is communicating with, we define  $\Omega^r(s) = \{T_t | T_t \in \Theta^r(s) \wedge T_s \in \Theta^r(t)\}$ . As before, we also define  $\Omega^r(s_1, \dots, s_k) = \Omega^r(s_1) \cap \dots \cap \Omega^r(s_k)$  and  $\Omega^r = \{T_1, \dots, T_{t(L)}\}$ .

Let  $N_s$  be the number of staff members of type  $T_s$ . The fraction of staff members being of type  $T_s$  equals  $n_s = N_s/N$ . The probability that  $m$  individuals who meet randomly can all use their preferred languages is then given by:

$$P_{rec}^m = \sum_{T_{s_1} \in \Omega^r} n_{s_1} \sum_{T_{s_2} \in \Omega^r(s_1)} n_{s_2} \sum_{T_{s_3} \in \Omega^r(s_1, s_2)} n_{s_3} \dots \sum_{T_{s_m} \in \Omega^r(s_1, s_2, \dots, s_{m-1})} n_{s_m} \tag{22}$$

## Appendix 2: Properties of $\rho$

It is easy to show that:

$$\rho = 1 - \frac{1}{L} \sum_{l=1}^L (1 - d_l)^2 = 1 - \frac{1}{L} \sum_{l=1}^L (1 - 2d_l - d_l^2) = \frac{2}{L} \sum_{l=1}^L d_l - \frac{1}{L} \sum_{l=1}^L d_l^2 \tag{23}$$

and therefore  $\rho = 2\mu - C/L$ , where  $C$  is equal to the Herfindahl–Hirschman concentration index (see formula 1).

To obtain a maximal polarization index under the condition  $D_1 + \dots + D_L = D_{max}$ , we consider the Lagrangian

$$\Lambda(D_1, \dots, D_L, \lambda) = 1 - \frac{1}{L} \sum_{i=1}^L \left(1 - \frac{D_i}{D}\right)^2 + \lambda \left(-D_{max} + \sum_{i=1}^L D_i\right) \tag{24}$$

Considering the partial derivatives of  $\Lambda$ , it is easy to see that the critical points have to satisfy  $D_1 = \dots = D_L$ .

## References

- Alesina, A., Baqir, R., & Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4), 1243–1284.
- Arcand, J.-L., & Grin, F. (2013). Language in economic development: Is English special and is linguistic fragmentation bad? In E. J. Erling & P. Seargeant (Eds.), *English and development. Policy, pedagogy and globalization* (pp. 243–266). Bristol: Multilingual Matters.
- Baldwin, K., & Huber, J. D. (2010). Economic versus cultural differences: Forms of ethnic diversity and public goods provision. *American Political Science Review*, 104(4), 644–662.
- Berthele, R., & Wittlin, G. (2013). Receptive multilingualism in the Swiss Army. *International Journal of Multilingualism*, 10(2), 181–195.
- Borbey, P., & Mendelsohn, M. (2017). *Le prochain niveau. Enraciner une culture de la dualité linguistique inclusive en milieu de travail au sein de la fonction publique fédérale*. Ottawa: Gouvernement du Canada.
- Bossert, W., D'Ambrosio, C., & La Ferrara, E. (2011). A generalized index of fractionalization. *Economica*, 78(312), 723–750.
- Bove, V., & Ruggeri, A. (2016). Kinds of blue: Diversity in UN peacekeeping missions and civilian protection. *British Journal of Political Science*, 46(3), 681–700.
- Braunmüller, K. (2013). Communication based on receptive multilingualism: Advantages and disadvantages. *International Journal of Multilingualism*, 10(2), 214–223.
- Cardinal, L. (2015). State tradition and language regime in Canada. In L. Cardinal & S. K. Sonntag (Eds.), *State traditions and language regimes* (pp. 29–43). Montreal: McGill-Queen's University Press.
- Christopher Guerra, S., & Zurbriggen, S. (2013). *Sprachkurse für Mitarbeitende der Bundesverwaltung. Evaluation und Analyse des Angebotes und dessen Nutzung. Ausführlicher Projektbericht*. Fribourg: Institut für Mehrsprachigkeit.
- Churchill, S. A., Ocloo, J. E., & Siawor-Robertson, D. (2017). Ethnic diversity and health outcomes. *Social Indicators Research*, 134(3), 1077–1112.
- Coray, R., Kobelt, E., Zwicky, R., Duchêne, A., & Kübler, D. (2015). *Mehrsprachigkeit verwalten? Spannungsfeld Personalrekrutierung beim Bund*. Zürich: Seismo.
- Desmet, K., Ortuno-Ortín, I., & Weber, S. (2009). Linguistic diversity and redistribution. *Journal of European Economic Association*, 7(6), 1291–1318.
- DFP. (2015). Promotion du plurilinguisme. Rapport d'évaluation au Conseil fédéral et recommandations sur la politique de plurilinguisme (art. 8d, al. 4, OLang). Développement de 2008 à 2014. Perspectives pour la période de 2015 à 2019. Berne. Département fédéral des finances (DFP), Déléguee fédérale au plurilinguisme (DFP).
- Ehmke-Gendron, S. (2015). *T2020: Report on steps taken and options for action*. Brussels: Directorate-General for Administration—Translation Service. Council of the European Union, General Secretariat.
- Esteban, J., Mayoral, L., & Ray, D. (2012). Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4), 1310–1342.
- Fall, P. L., & Zhang, Y. (2011). *Multilingualism in the United Nations system organisations: Status of implementation*. Geneva: Joint Inspection Unit - United Nations.
- Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2), 195–222.
- Fedderke, J., Luiz, J., & de Kadt, R. (2008). Using fractionalization indexes: Deriving methodological principles for growth studies from time series evidence. *Social Indicators Research*, 85, 257–278.
- García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Basingstoke/New York: Palgrave MacMillan.
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge: Cambridge University Press.
- Gaspard, H. (2015). Canada's official languages policy and the federal public service. In L. Cardinal & S. K. Sonntag (Eds.), *State traditions and language regimes* (pp. 191–204). Montreal: McGill-Queen's University Press.
- Gazzola, M. (2014). *The evaluation of language regimes: Theory and application to multilingual patent organisations*. Amsterdam/Philadelphia: John Benjamins.
- Gazzola, M. (2016a). Multilingual communication for whom? Language policy and fairness in the European Union. *European Union Politics*, 17(4), 546–569.
- Gazzola, M. (2016b). Programmazione e controllo della 'politica del plurilinguismo' nell'amministrazione federale svizzera. *Studi Italiani di Linguistica Teorica e Applicata*, 45(3), 479–497.
- General Assembly. (2017). *Multilingualism. Report of the Secretary-General, A/71/757*. New York: United Nations.
- Gini, C. (1912). *Variabilità e mutabilità. Contributo allo studio delle distribuzioni e delle relazioni statistiche*. Bologna: Tipografia di Paolo Cuppini.

- Ginsburgh, V., & Weber, S. (2005). Language disenfranchisement in the European Union. *Journal of Common Market Studies*, 43(2), 273–286.
- Ginsburgh, V., & Weber, S. (2016a). Linguistic distances and ethnolinguistic fractionalization and disenfranchisement indices. In V. Ginsburgh & S. Weber (Eds.), *The Palgrave handbook of economics and language* (pp. 137–174). Basingstoke: Palgrave.
- Ginsburgh, V., & Weber, S. (2016b). Linguistic diversity, standardization, and disenfranchisement: Measurements and consequences. In M. Gazzola & B.-A. Wickström (Eds.), *The economics of language policy* (pp. 95–140). Cambridge, MA: MIT Press.
- Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language*, 32(1), 109–115.
- Grin, F. (2006). Peut-on faire confiance au modèle '1+2'? Une évaluation critique des scénarios de communication dans l'Europe multilingue. *Revista de Llengua i Dret*, 45, 217–231.
- Grin, F. (2017). Translation and language policy in the dynamics of multilingualism. *International Journal of the Sociology of Language*, 243, 155–181.
- Grin, F., Amos, J., Faniko, K., Fürst, G., Lurin, J., & Schwob, I. (2015). *Suisse-Société multiculturelle*. Glarus/Chur: Rüeeggler/Somedia.
- Groupe T2020. (2016). Rapport du Groupe T2020 constitué par le Comité du personnel du Secrétariat général du Conseil de l'UE. Brussels: Council of the European Union. Inter alia services.
- Hall, T. E. (2013). Public participation in election management: The case of language minority voters. *American Review of Public Administration*, 33(4), 407–422.
- Ishizawa, H., & Stevens, G. (2007). Non-English language neighborhoods in Chicago, Illinois: 2000. *Social Science Research*, 36(6), 1042–1064.
- Kolisnichenko, N., & Rosenbaum, A. (2009). Building a new democracy in Ukraine: The unacknowledged issue of ethnic and linguistic diversity in public administration education and training. *Public Administration Review*, 69(5), 932–940.
- Kruse, J., & Ammon, U. (2018). The language planning and policy for the European Union and its failures. In C. S. K. Chua (Ed.), *Unintended language planning in a globalising world: Multiple levels of players at work* (pp. 39–56). Berlin: De Gruyter Open.
- Kübler, D., Kobelt, E., & Andrey, S. (2011). Vers une bureaucratie représentative. La promotion de la représentation et de la diversité linguistiques dans l'administration fédérale en Suisse et au Canada. *Canadian Journal of Political Science/Revue canadienne de science politique*, 44(4), 903–927.
- Kübler, D., & Zwicky, R. (2018). *Topkader und Mehrsprachigkeit in der Bundesverwaltung*, Studienberichte des Zentrums für Demokratie Aarau, Nr. 13. Aarau: Zentrums für Demokratie Aarau.
- Kudryavtsev, E., & Ouedraogo, L.-D. (2003). *Implementation of multilingualism in the United Nations system*. Geneva: Joint Inspection Unit.
- Laitin, D. (2000). What is a language community? *American Journal of Political Science*, 44(1), 142–155.
- Laitin, D., & Ramachandran, R. (2016). Language policy and human development. *American Political Science Review*, 110(3), 457–480.
- Lieberson, S. (1964). An extension of Greenberg's linguistic diversity measure. *Language*, 40(4), 526–531.
- Liu, A. H., & Pizzi, E. (2018). The language of economic growth: A new measure of linguistic heterogeneity. *British Journal of Political Science*, 48(4), 953–980.
- Lüdi, G. (2007). The Swiss model of plurilingual communication. In J. D. Ten Thije & L. Zeevaert (Eds.), *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts* (pp. 159–178). Amsterdam–Philadelphia: John Benjamins.
- McEntee-Atalians, L. (2015). Language policy and planning in international organisations. In U. Jessner-Schmid & C. Kramsch (Eds.), *The multilingual challenge: Cross-disciplinary perspectives* (pp. 295–322). Berlin: De Gruyter Mouton.
- Meylaerts, R., & Nuñez, G. G. (Eds.). (2017). *Translation and public policy: Interdisciplinary perspectives and case studies*. London: Routledge.
- Mondada, L., Markaki, V., Merlino, S., Oloff, F., & Traverso, V. (2013). Multilingual practices in professional settings: Keeping the delicate balance between progressivity and intersubjectivity. In A.-C. Berthoud, F. Grin, & G. Lüdi (Eds.), *Exploring the dynamics of multilingualism* (pp. 3–32). Amsterdam: John Benjamins.
- Naff, K. C., & Jurée Capers, K. (2014). The complexity of descriptive representation and bureaucracy: The case of South Africa. *International Public Management Journal*, 17(4), 515–539.
- Nettle, D. (2000). Linguistic fragmentation and the wealth of nations: The Fishman-Pool hypothesis reexamined. *Economic Development and Cultural Change*, 48(2), 335–348.
- Podestà, G. (2001). *Preparing for the Parliament of the Enlarged European Union*, PE 305.269/BUR/fin. Brussels: European Parliament.
- Rehbein, J., ten Thije, J., & Verschik, A. (2011). Lingua receptiva (LaRa)—Remarks on the quintessence of receptive multilingualism. *International Journal of Bilingualism*, 16(3), 248–264.

- Ritz, A., & Brewer, G. A. (2013). Does societal culture affect public service motivation? Evidence of sub-national differences in Switzerland. *International Public Management Journal*, 16(2), 224–251.
- Schaeffer, M. (2013). Can competing diversity indices inform us about why ethnic diversity erodes social cohesion? A test of five diversity indices in Germany. *Social Science Research*, 42(3), 755–774.
- Stirling, A. (2007). A general framework for analyzing diversity in science, technology and society. *Journal of the Royal Society, Interface*, 4, 707–719.
- Sturm, J.-E., & De Haan, J. (2015). Income inequality, capitalism, and ethno-linguistic fractionalization. *American Economic Review*, 105(5), 593–597.
- Ten Thije, J. D., & Zeevaert, L. (Eds.). (2007). *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts*. Amsterdam–Philadelphia: John Benjamins.
- Turgeon, L., & Gagnon, A. G. (2013). The politics of representative bureaucracy in multilingual states: A comparison of Belgium, Canada and Switzerland. *Regional & Federal Studies*, 23(4), 407–425.
- Turgeon, L., & Gagnon, A.-G. (2015). Bureaucratic language regimes in multilingual states: Comparing Belgium and Canada. In L. Cardinal & S. K. Sonntag (Eds.), *State traditions and language regimes* (pp. 119–136). Montreal: McGill-Queen's University Press.
- Van der Jeught, S. (2015). *EU language law*. Groningen: Europa Law Publishing.
- van Staveren, I., & Pervaiz, Z. (2017). Is it ethnic fractionalization or social exclusion, which affects social cohesion? *Social Indicators Research*, 130(2), 711–731.
- Voslamber, D. (2018). Choosing working languages in a multilingual organization. A statistical analysis with a particular view on the European Union. In M. Gazzola, T. Templin, & B.-A. Wickström (Eds.), *Language policy and linguistic justice: Economic, philosophical and sociolinguistic approaches* (pp. 337–360). Berlin: Springer.
- Werlen, I. (2007). Receptive multilingualism in Switzerland and the case Biel/Bienne. In K. Buhrig & J. Ten Thije (Eds.), *Beyond misunderstanding: Linguistic analyses of intercultural communication* (pp. 137–157). Amsterdam: John Benjamins.
- Wodak, R., Krzyżanowski, M., & Forchtner, B. (2012). The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions. *Language in Society*, 41(2), 157–186.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.